

G0971

CMOS WITH STRAINED SILICON CHANNEL NMOS AND SILICON GERMANIUM CHANNEL PMOS

TECHNICAL FIELD

The present invention relates to the field of manufacturing semiconductor devices, and more particularly, to an improved complementary metal oxide semiconductor field effect transistor (CMOSFET).

BACKGROUND OF THE INVENTION

An important aim of ongoing research in the semiconductor industry is increasing semiconductor performance while decreasing the size of semiconductor devices. Planar transistors, such as metal oxide semiconductor field effect transistors (MOSFET) are particularly well suited for use in high-density integrated circuits. As the size of the MOSFET decreases and the density of the devices increases, the demands of maintaining the electrical isolation of adjacent transistors increases.

The use of shallow trench isolation (STI) significantly shrinks the area needed to isolate transistors and thereby provides higher device density than local oxidation of silicon (LOCOS). STI also provides superior latch-up immunity, smaller channel width encroachment, and improved planarity. The use of STI techniques also eliminates the bird's beak frequently encountered with LOCOS.

Strained silicon technology allows the formation of higher speed devices. Strained-silicon transistors are created by depositing a graded layer of silicon germanium (SiGe) on a bulk silicon wafer. A thin layer of silicon is subsequently deposited on the SiGe layer. The distance between atoms in the SiGe crystal lattice is greater than the distance between atoms in an ordinary silicon lattice. Because there is a natural tendency of atoms inside different crystals to align with one another when a second crystal is formed over a first crystal, when silicon is deposited on top of SiGe the silicon crystal lattice tends to stretch or "strain" to align the silicon atoms with the atoms in the SiGe layer. Electrons in the strained silicon experience less resistance and flow up to 80% faster than in unstrained silicon.

There are two general types of MOS transistors, N-channel MOS (NMOS) formed with n-type source and drain regions in a p-type wafer, and P-channel MOS (PMOS) formed with p-type source and drain regions. NMOS transistors conduct electrons through the transistor channel, while PMOS

transistors conduct holes through the transistor channel. Typically, the source and drain regions of the transistors are doped with phosphorous or arsenic to form n-type source/drain regions, while boron doping is used to form p-type source/drain regions.

CMOS transistors, which comprise N- and P- channel MOS transistors on the same substrate, suffer from imbalance. The imbalance is due to electron mobility being greater than hole mobility in the channel region. Therefore, NMOS transistors are faster than PMOS transistors. Typically, NMOS transistors are about 2 to about 2.5 times faster than PMOS transistors.

The problem of imbalance in CMOS devices is further exacerbated in CMOS devices comprising strained silicon channels. Strained silicon does not enhance hole mobility in PMOS transistors as much as it does electron mobility in NMOS transistors. Therefore, a CMOS device comprising strained silicon channels is more unbalanced than a CMOS transistor with conventional crystalline silicon channels.

The term semiconductor devices, as used herein, is not to be limited to the specifically disclosed embodiments. Semiconductor devices, as used herein, include a wide variety of electronic devices including flip chips, flip chip/package assemblies, transistors, capacitors, microprocessors, random access memories, etc. In general, semiconductor devices refer to any electrical device comprising a semiconductor.

SUMMARY OF THE INVENTION

There exists a need in the semiconductor device art to provide CMOS transistors, which combine more closely balanced PMOS and NMOS transistors with the increased speed of strained silicon. There exists a need in this art to provide smaller devices. There exists a need in this art to combine higher speed and smaller device-size to provide a more balanced CMOS transistor.

These and other needs are met by embodiments of the present invention, which provide a semiconductor device comprising a semiconductor substrate comprising a silicon germanium (SiGe) layer formed on a base layer. A plurality of field effect transistors (FET) including at least one first FET and at least one second FET are formed on the semiconductor substrate. The first FET comprises a gate electrode formed over the SiGe layer, and a gate oxide layer interposed between the gate electrode and the SiGe layer, wherein the gate oxide layer is formed immediately adjacent to and in contact with the SiGe layer. The second FET comprises a silicon layer formed on the SiGe layer, a gate oxide layer formed on the silicon layer, and a gate electrode formed on the gate oxide layer.

The earlier stated needs are also met by certain embodiments of the instant invention, which provide a method of manufacturing semiconductor devices comprising providing a semiconductor substrate comprising a SiGe layer formed on a base layer and a silicon layer formed on the SiGe layer. The semiconductor substrate comprises first regions and second regions spaced apart from each other by interposed isolation regions. At least a portion of the silicon layer is selectively removed only in the first region. Dopant is implanted in the first and second regions. A gate oxide layer is formed in the first and second regions and a gate electrode layer is formed over the gate oxide layer.

This invention addresses the needs for a higher-speed CMOS device with improved balance between the PMOS transistor and the NMOS transistor. The present invention allows the formation of smaller-sized CMOS transistors with improved performance.

The foregoing and other features, aspects, and advantages of the present invention will become apparent in the following detailed description of the present invention when taken in conjunction with the accompanying drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

FIGS. 1-6 schematically illustrate the formation of a CMOS device according to one embodiment of the present invention.

FIGS. 7-8 schematically illustrate an alternative method of forming the gate oxide layer according to the present invention.

FIG. 9 illustrates the use of a low temperature, bottom anti-reflection coating to preserve the amorphous nature of amorphous polysilicon.

DETAILED DESCRIPTION OF THE INVENTION

The present invention enables the production of improved reduced-size, high-speed semiconductor devices. The present invention combines the enhanced electron mobility through NMOS transistors comprising strained silicon channels with the enhanced mobility of holes through PMOS transistors comprising SiGe channels. These benefits are provided by forming NMOS transistor channels comprising strained silicon and PMOS transistors without strained silicon in the channel region.

The invention will be described in conjunction with the formation of the semiconductor device in the accompanying drawings. However, this is exemplary only as the claimed invention is not limited to the formation of the specific device illustrated in the drawings.

The mobility of holes in a SiGe lattice is greater than the mobility of holes in either strained silicon or conventional silicon crystalline lattices. As discussed above, the mobility of electrons is greater in strained silicon than conventional crystalline silicon. Therefore, forming PMOS transistors with SiGe channels and NMOS transistors with strained silicon channels provides a more balanced CMOS device that operates at higher overall speed than a device in which the PMOS and NMOS channels comprise conventional silicon crystal lattices.

Semiconductor substrate 10, as shown in FIG. 1, comprises a base layer 12, such as a silicon wafer, with a SiGe layer 14 formed thereon. A layer of strained silicon 16 is formed over the SiGe layer 14. The base layer 12 is typically a silicon wafer about 100 μm thick. The SiGe layer 14 is formed by a chemical vapor deposition (CVD) process, such as ultra-high vacuum chemical vapor deposition (UHVCVD). The SiGe layer 14 comprises a sublayer 15, in which the concentration of Ge is graded from about 0% Ge at the SiGe layer 14/base layer 12 interface. The concentration is increased as the SiGe sublayer 15 is deposited up to a maximum concentration of about 30% Ge. In certain embodiments the maximum concentration of Ge is about 20%. The thickness of the graded sublayer 15 is about 2 μm . When the maximum desired concentration of Ge is achieved, about a 1 μm to 2 μm thick sublayer of SiGe 17, of substantially constant Ge concentration, is deposited over the graded sublayer of SiGe 15 forming a SiGe layer 14 with an overall thickness of about 3 μm to about 4 μm . The concentration of Ge in the constant concentration sublayer 17 is substantially the same as the maximum Ge concentration in the graded sublayer 15.

The strained silicon layer 16 is an epitaxial layer formed by CVD to a thickness of about 50 Å to about 500 Å. In certain embodiments, the silicon layer 16 thickness is about 100 Å to about 300 Å. The atoms in the lattice structure of silicon layer 16 stretch apart from each other in order to align themselves with the underlying lattice structure of the SiGe layer. Electron flow in this "strained silicon" layer 16 is much faster than in ordinary crystalline silicon.

A plurality of isolation regions 18 are formed in the semiconductor substrate 10 in order to electrically isolate a plurality of semiconductor devices formed in the semiconductor substrate 10. The isolation regions 18 illustrated in FIG. 1 are shallow trench isolation (STI) regions formed by anisotropically etching semiconductor substrate 10 to form trenches. A thermal oxide liner layer 20 is grown by conventional methods to a thickness of about 30 Å to about 100 Å, such as by exposing the

semiconductor substrate 10 to an oxygen ambient at a temperature about 850° C to about 1150° C. The trenches are subsequently filled with an insulating material 22, such as silicon nitride or silicon oxide, by a conventional CVD process to form the isolation region 18. Some of the conventional methods of filling a trench with silicon oxide include a) tetraethylorthosilicate low pressure chemical vapor deposition (TEOS LPCVD), b) non-surface sensitive TEOS-ozone atmospheric pressure or sub-atmospheric chemical vapor deposition (APCVD or SACVD), and c) silane oxidation high-density plasma CVD.

The relatively high aspect ratio of the STI trench raises the concern of incomplete filling or void formation when the trench is filled with insulating material 22. In order to alleviate void formation, the insulating material 22 can be densified by thermally annealing the filled trench at high temperature.

A mask 24 is formed over semiconductor substrate 10 with an opening 26 at the location where the PMOS transistor is to be formed, as shown in FIG. 2. Subsequent to forming the mask 24, the exposed silicon layer 16 is removed, such as by anisotropic etching, and a PMOS channel ion implant is performed. The same mask 24 used for etching the exposed silicon layer 16 is also used during the PMOS channel ion implant. The PMOS channel ion implant is performed through the mask opening 26, either before or after the removal of the exposed silicon layer 16. After removing the strained silicon layer where the PMOS will be formed, two distinct regions are formed, as shown in FIG. 3; a first region 28, in which the surface 13 of the SiGe layer 14 is exposed, and a second region 30, in which the strained silicon layer 16 is exposed.

In a step not shown, the mask 24 is removed and another mask is formed with an opening over the second region 30 where the NMOS is to be formed, and a NMOS channel ion implant is performed while the first region 28 is masked. However, in this step, the strained silicon layer 16 is not etched in the second region 30.

In certain embodiments of the instant invention, a mask is first formed with an opening over the region where the NMOS is to be formed, a NMOS channel implant is performed, and the mask is removed. Following the NMOS channel implant, a second mask is formed with an opening over the region where the PMOS is to be formed, and then the strained silicon layer removal and the PMOS channel implant is performed.

After performing channel implantation in the first and second regions 28, 30, the transistor gates are formed. A gate oxide layer 32 and a gate electrode layer 34 are formed on semiconductor

substrate 10, as shown in FIG. 4. The gate oxide layer 32 is formed to a thickness of about 10 Å to about 100 Å, by either CVD or thermal oxidation. In certain embodiments, the gate oxide layer is formed to a thickness of about 10 Å to about 50 Å. The gate oxide layer 32 can be a deposited silicon oxide or a high-k dielectric, such as zirconium dioxide, hafnium dioxide, or stacked nitride dielectric. The gate electrode layer 34 is formed by depositing polysilicon to a thickness of about 50 nm to about 500 nm. In certain embodiments, the gate electrode layer thickness is about 100 nm to about 300 nm.

A patterned mask is formed over the structure of FIG. 4, such as by photolithographic patterning, the gate oxide layer 32 and the gate electrode layer 34 are etched to form gate electrode structures, and ion implantation is performed to form source/drain extensions 36, 38, as shown in FIG. 5. Separate masking and ion implantation steps are carried out in order to form source/drain extensions in the first region 28 and the second region 30 because different types of ions are implanted in each region. For example, the second region 30 is masked while ions are implanted in the first region 28, and conversely, the first region 28 is masked while ions are implanted in the second region 30. As shown in FIG. 5, source/drain extensions 38 are formed in the first region 28 and source/drain extensions 36 are formed in the second region 30.

Sidewall spacers 40, as shown in FIG. 6, are subsequently formed on the semiconductor substrate 10 by depositing a layer of insulating material, such as silicon nitride or silicon oxide followed by anisotropic etching to form sidewall spacers 40. PMOS source/drain regions 44 and NMOS source/drain regions 42 are subsequently formed by conventional ion implantation techniques, followed by annealing to form the respective p-type and n-type source/drain regions 44, 42 comprising the respective lightly doped drain extensions 38, 36 and heavily doped regions 45, 43. Separate masking and ion implantation steps are carried out in order to form the heavily doped regions 45, 43 of the source/drain regions 44, 42 in the first region 28 and the second region 30 because different types of ions are implanted in each region. As shown in FIG. 6, the resulting structure is a CMOS device with PMOS 46 and NMOS 48 transistors, wherein the channel region of the PMOS comprises the SiGe layer 14 and the channel region of NMOS 48 comprises the strained silicon layer 16 and the SiGe layer 14.

In an alternative embodiment of forming the CMOS of the present invention, the silicon layer 16 is only partially removed in the first region 28. As shown in FIG. 7, the silicon layer 16 in first region 28 is not completely removed, rather it is etched down to a thickness of about 5 Å to about 25 Å by anisotropic etching. A gate oxide layer 50 is formed over the first and second regions 28, 30 by thermally oxidizing the strained silicon layer 16. The thermal oxidation substantially completely oxidizes the silicon layer 16 remaining in the first portion 28, while only partially oxidizing silicon

layer 16 in second region 30. To provide a gate oxide layer 50 of about 10 Å thickness, a silicon layer thickness of about 5 Å is required in first region 28 before thermal oxidation. To provide a gate oxide layer 50 thickness of about 20 Å, the thickness of the remaining silicon layer in the first region 28 before thermal oxidation needs to be about 10 Å. After forming the thermal gate oxide layer 50 the semiconductor substrate 10 is processed as described above, including the steps of forming a gate electrode layer, masking and etching the gate electrode layer and gate oxide layer, forming sidewall spacers, and ion implantation to form a CMOS device.

In other aspects, a low temperature plasma enhanced chemical vapor deposition (PECVD) bottom anti-reflection coating (BARC) is spun on an amorphous layer of polysilicon. The use of a low temperature PECVD coating preserves the amorphous nature of the polysilicon layer, which helps maintain a smooth polysilicon line edge. Conventional BARC layers deposited at high temperature causes crystallization of the amorphous polysilicon, which results in rough polysilicon line edges. FIG. 9 illustrates a semiconductor substrate 60 comprising a low temperature BARC 66 formed on an amorphous polysilicon layer 64 on a silicon base layer 62, with a resist layer 68 overlying the BARC 66.

The CMOS device of the present invention is a reduced dimension, high-speed, more balanced CMOS device than CMOS devices formed on conventional crystalline silicon lattices. The methods of the present invention combine the enhanced mobility of holes in SiGe lattices and the enhanced mobility of electrons in strained silicon lattices to produce a higher-speed, more-balanced CMOS device.

The embodiments demonstrated in the instant disclosure are for illustrative purposes only. It should not be construed to limit the scope of the claims. As is clear to one of ordinary skill in the art, the instant disclosure encompasses a wide variety of embodiments not specifically illustrated herein.